

11-2014

Bayesian Inference for Volatility of Stock Prices

Juliet G. D'Cunha

Mangalore University, Mangalagangothri, Karnataka, India, gratiajuliet@gmail.com

K. A. Rao

Mangalore University, Mangalagangothri, Karnataka, India, arunaraomu@gmail.com

 Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

Recommended Citation

D'Cunha, Juliet G. and Rao, K. A. (2014) "Bayesian Inference for Volatility of Stock Prices," *Journal of Modern Applied Statistical Methods*: Vol. 13 : Iss. 2 , Article 29.

DOI: 10.22237/jmasm/1414816080

Emerging Scholars: Bayesian Inference for Volatility of Stock Prices

Juliet Gratia D’Cunha
Mangalore University
Mangalagangothri, Karnataka, India

K. Aruna Rao
Mangalore University
Mangalagangothri, Karnataka, India

Lognormal distribution is widely used in the analysis of failure time data and stock prices. Maximum likelihood and Bayes estimator of the coefficient of variation of lognormal distribution along with confidence/credible intervals are developed. The utility of Bayes procedure is illustrated by analyzing prices of selected stocks.

Keywords: Bayesian inference, volatility, stock prices, coefficient of variation, lognormal distribution

Introduction

The study on coefficient of variation (CV) of the normal distribution dates back to McKay (1932); since then various articles have appeared concerning improved estimation of CV of a normal distribution and tests for equality of CV’s of two or more normal distributions. Some of the recent references regarding the estimation of CV of the normal distribution are Ahmed (1995), Breunig (2001), Liu, et al. (2006), Mohmoudvand & Hassani (2009) and Panichkitkosolkul (2009). The papers dealing with tests for equality of CV’s of independent normal distributions are Bennett (1976), Doornabos & Dijkstra (1983), Shafer & Sullivan (1986), Gupta & Ma (1996), Nairy & Rao (2003) and Verril & Johnson (2007). In addition to these papers, the papers on CV relating to finance and economics are Brief & Owen (1969), Jobson & Korkie (1981), De, et al. (1996) and Memmel (2003). These papers are developed on the assumption of normality of the observations.

Juliet Gratia D’Cunha is a Research Scholar in the Department of Statistics. Email her at gratiajuliet@gmail.com. Dr. Rao is a Professor in the Department of Statistics. Email him at arunaraomu@gmail.com.

BAYESIAN INFERENCE FOR VOLATILITY OF STOCK PRICES

Generally stock prices do not follow normal distribution and the data is analyzed using logarithm of prices. This amounts to the assumption that stock price is lognormally distributed. CV is not invariant under distributional transformation, and thus estimators are to be derived for the CV of the lognormal distribution.

Maximum likelihood estimator (M.L.E) and confidence interval for the CV of the lognormal distribution are derived, as well as the Bayes estimator of CV of the lognormal distribution using a) Right invariant prior b) Left invariant Jeffrey's prior.

Bayesian inference has several advantages over the likelihood based inference (Ghosh, et al., 2006; Berger, 1985). Simulation study carried out in this paper suggests that Bayesian credible intervals have smaller average length compared to the confidence interval obtained by M.L.E. Financial analysts are generally not well exposed to Bayesian analysis and this paper introduces this idea by analyzing the stock prices of 3 Indian stocks.

The maximum likelihood estimator and Bayes estimator of the CV of the lognormal distribution and the associated confidence/credible intervals are initially derived. A simulation study is conducted to compare the coverage probability and average length of the confidence/credible intervals. The procedures developed in this paper are illustrated by analyzing stock prices of 3 scripts belonging to large cap sector of the Indian stock market. For this purpose daily data from August 19 to November 6 for the year 2013 is used. By using part of the data as training set and remaining data as the validation set, the paper demonstrates that Bayesian inference can be used to predict stock market volatility.

Bayes Estimator of CV of the Lognormal Distribution

Let x_1, x_2, \dots, x_n be a random sample from lognormal distribution with density

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{(\log x - \mu)^2}{2\sigma^2}}, x > 0, -\infty < \mu < \infty, \sigma > 0 \quad (1)$$

Denoting $\log X_i$ as Z_i , the minimal sufficient statistic for μ and σ^2 are

$$\bar{Z} = \frac{\sum_{i=1}^n Z_i}{n} \quad (2)$$

and

$$S_z^2 = \frac{\sum_{i=1}^n (Z_i - \bar{Z})^2}{n} \quad (3)$$

Therefore the maximum likelihood estimator of μ and σ^2 are \bar{Z} and $\left(\frac{n-1}{n}\right)S_z^2$. The mean, variance and coefficient of variation of the lognormal distribution are

$$E(x) = e^{\mu + \frac{1}{2}\sigma^2} \quad (4)$$

$$V(x) = \left(e^{2\mu + \sigma^2}\right)\left(e^{\sigma^2} - 1\right) \quad (5)$$

$$CV = \left(e^{\sigma^2} - 1\right)^{\frac{1}{2}} \quad (6)$$

respectively. Using the invariance property of maximum likelihood estimators, the maximum likelihood estimator of the CV of lognormal distribution is given by

$$\left(e^{S_z^2} - 1\right)^{\frac{1}{2}}, \text{ (Calculations herein used } \frac{n-1}{n} \approx 1) \quad (7)$$

The Bayes estimator of the CV of the lognormal distribution depends on the specification of the prior distribution for μ and σ^2 . In objective Bayesian analysis, the commonly used priors are the following

- *Right invariant prior:* For the location scale family with location parameter μ and scale parameter σ , the right invariant prior is $\pi(\mu, \sigma) = 1/\sigma$.
- *Jeffrey's prior:* Jeffrey's prior for μ and σ is given by $\pi(\mu, \sigma) = 1/\sigma^2$. Jeffrey's prior is left invariant but not right invariant.

Because the lognormal distribution belongs to log location scale family, the above priors were used in this study. Although right invariant prior is recommended (Ghosh, et al., 2006; Berger, 1985), the use of Jeffrey's prior aids in studying the

BAYESIAN INFERENCE FOR VOLATILITY OF STOCK PRICES

Bayesian robustness with respect to specification of the prior distribution. Because the distribution of \bar{Z} and S_z^2 are independent, denoting $\eta = 1/\sigma^2$, after some simplification the posterior density of η is obtained as Gamma $\left(\frac{n+2}{2}, \frac{1}{2}(n-1)S_z^2\right)$ is obtained under right invariant prior and Gamma $\left(\frac{n+3}{2}, \frac{1}{2}(n-1)S_z^2\right)$ under Jeffrey's prior.

Under squared error loss function, the Bayes estimator of CV is

$$E\left(\left(e^{\sigma^2} - 1\right)^{\frac{1}{2}}\right) = E\left(\left(e^{\frac{1}{\eta}} - 1\right)^{\frac{1}{2}}\right) \quad (8)$$

where the expectation is taken with respect to the posterior density of $\pi(\eta|z)$. This expectation must be evaluated numerically, thus the importance sampling approach was used to evaluate the integral. In this approach observations are generated from the posterior density and the numerical value of the expectation is given by

$$E\left(e^{\sigma^2} - 1\right)^{\frac{1}{2}} = \frac{1}{M} \sum_{i=1}^M \left(e^{\sigma_i^2} - 1\right)^{\frac{1}{2}} \quad (9)$$

where $\sigma_i^2, i = 1$ to M refers to the value of $1/\eta_i$ generated from the posterior density and M denotes the number of sample values generated. 10,000 observations are generated from the posterior density and using this, the Bayes estimator and equi-tailed credible intervals are obtained. For the likelihood based confidence interval, the equi-tailed confidence interval for $\eta = 1/\sigma^2$ is constructed using the Chi-square distribution for $\frac{(n-1)}{\sigma^2} S_z^2$. This confidence interval is then inverted to give a confidence interval for CV of the lognormal distribution. The confidence interval based on maximum likelihood estimator is given by

$$\left[\left(e^{\sigma_L^2} - 1\right)^{\frac{1}{2}}, \left(e^{\sigma_U^2} - 1\right)^{\frac{1}{2}} \right] \quad (10)$$

where

$$\sigma^2_L = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\chi^2_{\frac{\alpha}{2}}(n-1)} \quad (11)$$

and

$$\sigma^2_U = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\chi^2_{(1-\frac{\alpha}{2})}(n-1)} \quad (12)$$

Finite Sample Comparison of Credible and Confidence Intervals

The advantage of Bayes inference over likelihood-based inference is that it gives straightforward interpretation of the credible interval. Nevertheless, the superiority of the Bayes inference follows by comparing the coverage probability and length of the credible interval compared to the confidence interval based on maximum likelihood estimator. For this purpose a simulation study is conducted. For a random sample of size n ($n = 10, 20, 40, 60, 80, 100, 150, 200$, observations are generated from lognormal distribution or equivalently from normal distribution) with parameter μ and σ^2 . The value of μ and σ^2 are adjusted to yield a CV of 0.1, 0.3, 0.5, 0.7, 1, 1.5, 2, 2.5. The value of μ is fixed at 3. For the sample size and the value of CV, maximum likelihood estimator and the associated confidence intervals are computed using the expressions given in the previous section. For this sample size and value of CV, Bayes estimator, equi-tailed and HPD credible intervals are obtained using 10,000 simulated values of η , and thereby $(e^{\frac{1}{\eta}} - 1)^{\frac{1}{2}}$ from the posterior gamma density of η . This constitutes a single run in the simulation experiment. In each run the length of the confidence/credible interval is recorded. In addition, it is also recorded that whether the true value lies inside the confidence/credible interval. To estimate the coverage probability and average length of the confidence interval, the simulation experiment is repeated using 1000 runs. The coverage probability refers to the proportion of times the true value lies inside the interval. The credible/confidence level is fixed at 0.95. Tables 1 and 2 summarize the results of the simulation study.

BAYESIAN INFERENCE FOR VOLATILITY OF STOCK PRICES

Table 1. Coverage probability of the credible and confidence interval for the CV across sample sizes for 8 combinations of specified values of CV

Sample Size	Bayes Procedure (Equi-tailed)				Maximum Likelihood (Equi-tailed)	
	# of times Coverage probability Is maintained		Average length		# of times Coverage probability Is maintained	Average length
	Right invariant prior	Jeffrey's prior	Right invariant prior	Jeffrey's prior		
10	0	0	*	*	8	19.0641
20	0	0	*	*	8	2.4722
40	0	0	*	*	8	1.1390
60	1	0	1.4965	*	8	0.8264
80	4	0	0.1812	*	8	0.6888
100	8	0	0.5513	*	8	0.5976
150	8	7	0.4472	0.5010	8	0.4715
200	7	5	0.4363	0.4342	7	0.4477
Overall	28	12	0.6225	0.4676	63	3.2134

* Whenever coverage probability is not maintained average length has not been calculated

It may be said that the coverage probability is maintained if the estimated coverage probability lies between 0.940 to 0.960. That is $(1-\alpha) \pm 0.01$. From the table it is clear that the confidence interval based on maximum likelihood estimator maintains coverage probability for all sample sizes. On the other hand the equi-tailed credible interval maintains coverage probability when the sample size is greater than or equal to 100. However the average length of the credible interval is much shorter compared to the confidence interval. For example when $n = 150$ using right invariant prior, the average length of the credible interval is 0.4472 and using Jeffrey's prior it is 0.5010 while for the confidence interval it is 0.4715. The average length of the interval is computed using those intervals for which the coverage probability is maintained. The length of the confidence interval for Jeffrey's prior is marginally higher than right invariant prior. Table 2 presents the coverage probability and length of the HPD credible interval.

Table 2 shows that HPD credible interval maintains coverage probability when the sample size is greater than or equal to 40. The average length of the HPD credible intervals for both right and left invariant priors is marginally larger than the equi-tailed credible intervals. Theoretically the length of the HPD credible interval should be shorter than equi-tailed credible interval. To explore the reason for this phenomenon the posterior density for sample size $n = 60$ and 100 were plotted and the histogram and frequency curve of the simulated distribution of $\left(e^{\frac{1}{n}} - 1\right)^{\frac{1}{2}}$ was also plotted.

Table 2. Coverage probability of the HPD credible interval for the CV across sample sizes for 8 combinations of specified values of CV

Sample Size	Highest Posterior Density (HPD)			
	# of times Coverage probability Is maintained		Average length	
	<i>Right invariant prior</i>	<i>Jeffrey’s prior</i>	<i>Right invariant prior</i>	<i>Jeffrey’s prior</i>
10	0	0	*	*
20	0	0	*	*
40	7	0	0.8344	*
60	7	2	0.7933	0.8164
80	6	6	0.3071	0.3009
100	8	8	0.5684	0.5563
150	8	7	0.4562	0.5109
200	8	7	0.3899	0.4382
Overall	44	30	0.5582	0.5244

* Whenever coverage probability is not maintained average length has not been calculated

The posterior density of η is gamma and thus the plot of the density function is smooth. From the histogram and frequency curve it becomes clear that the frequency curve needs to be smoothened at the tail areas. This type of smoothing does not affect the length of the HPD credible interval, but increases the length of the equi-tailed credible interval. This is the reason why the equi-tailed credible intervals are marginally shorter than the HPD credible interval. To incorporate any type of smoothing of a frequency curve in a simulation study is computationally prohibitive and is not attempted here. Figures 1 to 4 represent the posterior density of η and the histogram obtained from 10,000 simulated values of the distribution of $(e^{\sigma^2} - 1)^{\frac{1}{2}}$, corresponding to $n = 60$ and 100, for left and right invariant priors and the value of S_z^2 is fixed at 0.0862 for CV=0.3.

An attempt is also made to study the effect of specified value of CV on the length of credible/confidence interval. Table 3 presents the average length of the interval for various values of CV. From the table it becomes clear that the average length increases as the CV increases for the credible/confidence intervals. The length of the credible interval for the sample size n=100, a large value of CV=2.5, for HPD credible interval using right invariant prior is 1.7358 and using Jeffrey’s prior is 1.6924 and for confidence interval it is 1.8445. For equi-tailed credible interval for right invariant and Jeffrey’s prior it is 1.6747 and 1.6338. The difference in the average length of the confidence interval when CV=0.1 and 2.5, is minimum for equi-tailed credible interval using Jeffrey’s prior and is maximum for confidence interval based on M.L.E. The difference in average length for the HPD credible

BAYESIAN INFERENCE FOR VOLATILITY OF STOCK PRICES

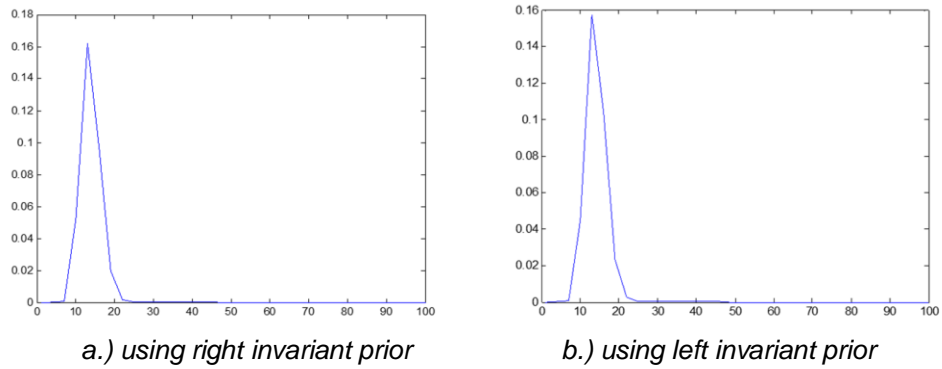


Figure 1. Posterior density of η when $n = 60$

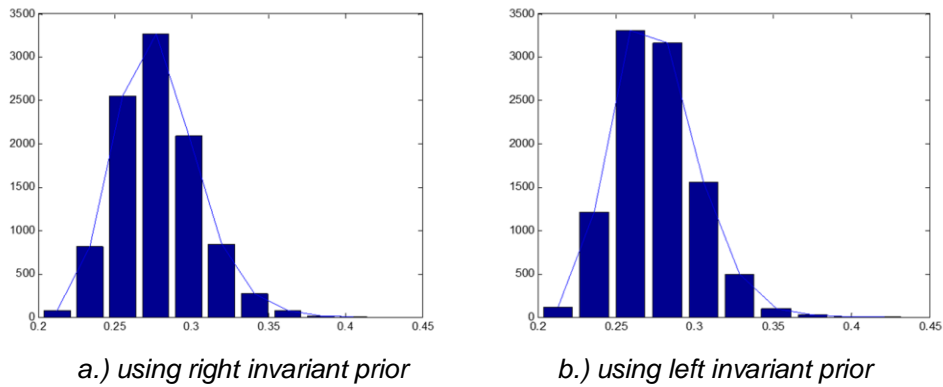


Figure 2. Histogram for $(e^{(1/\eta)} - 1)^{1/2}$ for $n = 60$

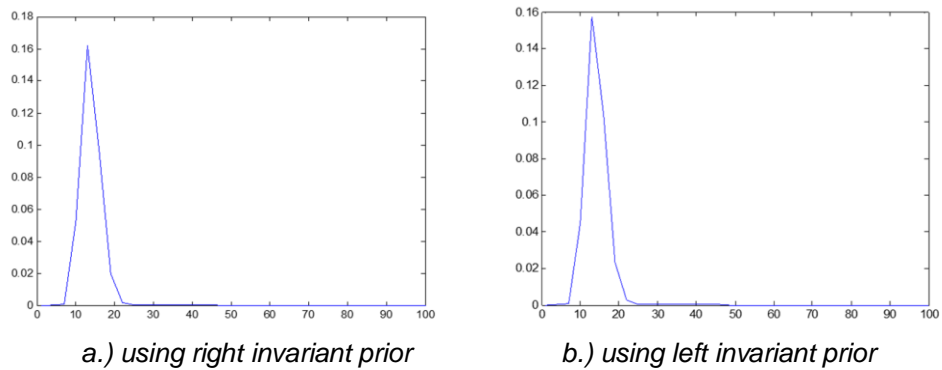


Figure 3. Posterior density of η when $n = 100$

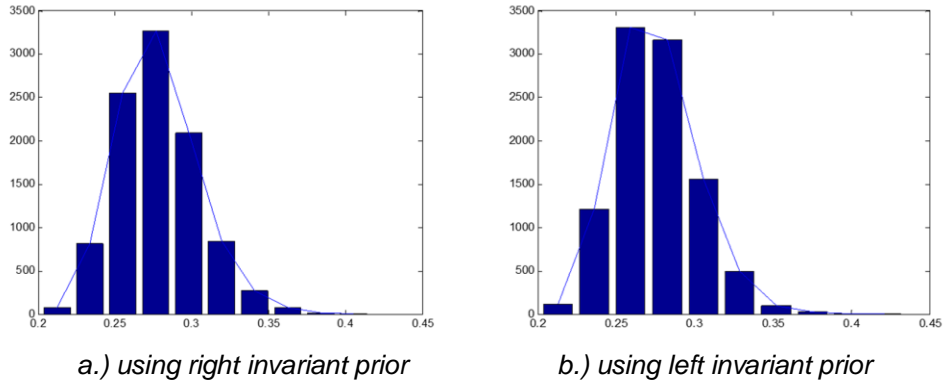


Figure 4. Histogram for $(e^{(1/n)}-1)^{1/2}$ for $n = 100$

interval based on right invariant and left invariant priors are 1.7080 and 1.6649. The same pattern can be observed for other sample sizes. The average length of HPD credible interval for Jeffrey’s prior is marginally higher compared to right invariant prior for all sample sizes and all values of CV under consideration. The coverage probability for these two priors indicates that the coverage probabilities are nearly the same. From the objective Bayesian analysis it amounts to the fact that Bayes procedure is robust against the specification of right and left invariant priors.

Table 3. Average length of the credible and confidence intervals for various values of CV when the sample size is $n = 100$.

Type of interval	Average length when CV equal to							Range	
	0.1	0.3	0.5	0.7	1	1.5	2		2.5
Equi-tailed credible interval with right invariant prior	0.0274	0.0858	0.1527	0.2322	0.3848	0.7153	1.1378	1.6747	1.6473
Equi-tailed credible interval with left invariant prior	0.0271	0.0848	0.1508	0.2337	0.3876	0.7205	1.1138	1.6338	1.6067
Confidence interval based on M.L.E	0.0284	0.0891	0.1593	0.2438	0.4077	0.7689	1.2393	1.8445	1.8161
HPD credible interval with right invariant prior	0.0278	0.0872	0.1554	0.2368	0.3937	0.7354	1.1748	1.7358	1.7080
HPD credible interval with left invariant prior	0.0275	0.0862	0.1534	0.2337	0.3876	0.7205	1.1489	1.6924	1.6649

Analysis of Stock Prices

The advantage of Bayesian analysis is that one can constantly upgrade their knowledge regarding the parameter. This is helpful for making future prediction. In this example the Bayes estimation of the index volatility per mean return is discussed with respect to the stock prices of 3 scripts belonging to large cap category, namely RELIANCE, ACC and TATASTEEL, of the Indian stock market. The daily data from August 19 to November 6, 2013 is used in this analysis. Starting with one week daily data as the training set, Bayes credible interval is obtained for the volatility per mean return. Subsequently the Bayes estimator for successive weeks is computed and the process is continued till the week for which the Bayes estimator lies outside the credible interval. The exercise is repeated with various starting weeks. Table 4 summarizes these results.

Table 4. Bayes credible interval for the index volatility per mean return based on 1 week data and the Bayes estimator for the successive weeks for different starting values.

Stock	Starting Value	95% credible interval	Bayes Estimator		
			2 nd week	3 rd week	4 th week
RELIANCE	Sept 17 th -Sept 23 rd	[0.0877,0.2714]	0.1507	0.1486	0.1460
ACC		[0.0396,0.1226]	0.0987	0.1219	0.1425
TATASTEEL		[0.0124,0.0384]	0.0265	0.0821	0.0820
RELIANCE	Oct 1 st - Oct 8 th	[0.0865,0.2745]	0.1460	0.1421	0.1103
ACC		[0.0713,0.2265]	0.1425	0.1570	0.1704
TATASTEEL		[0.0482,0.1486]	0.0820	0.1381	0.0164

Table 4 shows that based on one week data, the index for the subsequent week for all the three stocks can be accurately predicted. This is true regardless of the starting date namely August 19, September 17, October 1, etc. The duration of the data for making future predictions was also examined. For this purpose credible intervals were constructed using the first 2 through 10 weeks of data. To save space the results are not reported here. From these results it follows that by increasing the length of the data one do not get much accurate prediction for the successive week. Therefore it may be concluded that minimum data of one week be used for making prediction regarding volatility of the stock prices. If the duration increases, then the volatility increases thereby decreasing the decision of the future forecast.

Subjective Bayesian Analysis

As pointed out previously the advantage of Bayesian analysis is that the decision maker can use his belief for making future prediction. In the present scenario this can be achieved using conjugate prior. In the case of lognormal distribution, the conjugate prior is gamma for the scale parameter $\eta=1/\sigma^2$ where μ is fixed. Thus using Uniform prior for μ , the posterior distribution turns out to be gamma and one can use the program developed in this paper for carrying out subjective Bayesian analysis. The mean and variance of the posterior gamma density is given by $\alpha\beta$ and $\alpha\beta^2$ where $\alpha = (n+2)/2$ and $\beta = 1/2(n-1)S_z^2$ under right invariant prior. The parameters α and β can be determined by using past information as well as the subjective belief of the decision maker. The posterior density of the previous week can be used as the prior density for the week under consideration. In addition, the investigator can use his belief to modify the parameters of the posterior density of the previous week. Using past data, this type of subjective Bayesian analysis cannot be carried out and is not attempted in this paper.

Conclusion

This paper concentrates on the Bayesian estimation of the index, namely volatility per mean return. This is a frequently used indicator in the analysis of stock market data. The investigation indicates that Bayes credible intervals have smaller width compared to the confidence interval based on maximum likelihood estimator. Frequentist comparison of the credible interval and confidence interval in terms of coverage probability is not well accepted among the Bayesians. The results of this study support the view that accurate prediction can be made based on a small sample size of $n = 5$ for the volatility per mean return of stock prices. Caution has to be exercised for interpreting the width of the credible/confidence interval. For example if the width increases or decreases by 0.05, this amounts to a percentage change of 25% when $CV = 0.2$. Therefore one should not conclude that the difference in the average length of the credible interval and confidence interval is only marginal. The purpose of this paper is to demonstrate the utility of Bayesian inference for forecasting the stock prices.

This paper derives Bayes estimator and the associated credible intervals for the CV of the lognormal distribution. Lognormal distribution has applications in many areas like reliability studies and survival analysis where the focus is the duration of the lifetime. Although emphasis is given to the estimation of mean and median lifetime, the effectiveness of any treatment regime lies in the control of

BAYESIAN INFERENCE FOR VOLATILITY OF STOCK PRICES

variability in duration of lifetime. The results developed in this paper can also be used by researchers in these areas. Lognormal distribution is also used in the analysis of rainfall data (Ananthakrishnan & Soman, 1989) and the primary concern is the variability in rainfall, which is commonly measured using coefficient of variation. In these areas the data can be analyzed using objective Bayesian analysis of CV developed in this paper. Numerical analysis is carried out by writing programs using MATLAB software version 7.0 and can be obtained from the first author.

Acknowledgements

The first author would like to thank Government of India, Ministry of Science and Technology, Department of Science and Technology, New Delhi, for sponsoring her with an INSPIRE fellowship, which enables her to carry out the research program which she has undertaken. She is much honored to be the recipient of this award.

References

- Ahmed, S. E. (1995). A pooling methodology for coefficient of variation. *Sankya, Series B*, 57(1), 57-75.
- Ananthakrishnan, R., & Soman, M. K. (1989). Statistical distribution of daily rainfall and its association with the co-efficient of variation of rainfall series. *International Journal of Climatology*, 9, 485-500.
- Bennett, B. M. (1976). On an approximate test for homogeneity of coefficients of variation. In: W. J. Ziegler, (Ed.), *Contribution to Applied Statistics*, (169-171). Stuttgart: Birkhäuser.
- Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis* (2nd Ed.). New York: Springer-Verlag.
- Breunig, R. (2001). An almost unbiased estimator of the co-efficient of variation. *Economics Letters*, 70(1), 15-19.
- Brief, R. P., & Owen, J. (1969). A note on earnings risk and the co-efficient of Variation. *Journal of Finance*, 24, 901-904.
- De, P., Ghosh, J. B., & Wells, C. E. (1996). Scheduling to minimize the co-efficient of variation. *International Journal of Production Economics*, 44, 249-253.

Doornbos, R., & Dijkstra, J. B. (1983). A multi sample test for the equality of co-efficient of variation in normal population. *Communication in Statistics – Simulation and Computation*, 12, 147-158.

Ghosh, J. K., Delampady, M., & Samanta, T. (2006). *An introduction to Bayesian analysis: Theory and Methods*. New York: Springer.

Gupta, C. R., & Ma, S. (1996). Testing the equality of co-efficient of variation in 'k' normal populations. *Communication in Statistics – Theory and Methods*, 25, 115-132.

Jobson, J. D., & Korkie, B. M. (1981). Performance hypothesis testing with the sharpe and treynor measures. *Journal of Finance*, 36, 889-908.

Liu, W., Pang, W. K., & Huang, W. K. (2006). Exact confidence bounds for the co-efficient of variation of a normally distributed population. *International Journal of Statistics and Systems*, 1(1), 81-86.

McKay, A. T. (1932). Distribution of the co-efficient of variation and the extended *t*-distribution. *Journal of the Royal Statistical Society*, 95, 696-698.

Memmel, C. (2003). Performance hypothesis testing with the Sharpe ratio. *Finance Letters*, 1, 21-23.

Mohmoudvand, R., & Hassani, H. (2009). Two new confidence intervals for the CV in a normal distribution. *Journal of Applied Statistics*, 36(4), 429-442.

Nairy, K. S., & Rao, K. A. (2003). Tests of coefficient of variation in normal population. *Communication in Statistics – Simulation and Computation*, 32(3), 641-661.

Panichkitkosolkul, W. (2009). Improved confidence intervals for a co-efficient of variation of a normal distribution. *Thailand Statistician*, 7(2), 193-199.

Shafer, N. J. & Sullivan, J. A. (1986). A simulation study of a test for the equality of co-efficient of variation. *Communication in Statistics – Simulation and Computation*, 15, 681-695.

Verrill, S. P., Johnson, R. A., & Forest Products Laboratory (U.S.). (2007). *Confidence bounds and hypothesis tests for normal distribution coefficients of variation*. Madison, WI: U.S. Dept. of Agriculture, Forest Service, Forest Products Laboratory.